UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/621,474	07/17/2003	Hiroshi Nomiyama	JP920020109US1	2462

25259 7590 09/11/2006

IBM CORPORATION
3039 CORNWALLIS RD.
DEPT. T81 / B503, PO BOX 12195
REASEARCH TRIANGLE PARK, NC 27709

EXAMINER	
ROSE, HELENE ROBERTA	

ART UNIT	PAPER NUMBER
2163	

DATE MAILED: 09/11/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/621,474	NOMIYAMA ET AL.	
	Examiner	Art Unit	
	Helene Rose	2163	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *28 August 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *8,10-12 and 15-20* is/are pending in the application.

4a) Of the above claim(s) *1-7,9,13 and 14* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *8,10-12 and 15-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *17 July 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

Detailed Action

1. In response to communications filed on April 04, 2006, Claims 8 and 10-13 were amended; Claims 1-7,9 and 14 were cancelled and new claims 15-20 were added per applicant request.

2. This is a **RESPONSE** to the **AMENDMENT AFTER FINAL** filed on 8/28/2006 in which Claim 13 has been cancelled; No claims were added nor amended; Therefore, Claims 8,10-12 and 15-20 are still pending.

3. Applicant's request for reconsideration of the finality of the rejection of the last Office action is persuasive and, therefore, the finality of that action is withdrawn.

Claim Rejections – 35 U.S.C 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. Claims 8,10-13, and15-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ishikawa et al (US Patent No. 5,848,407, Date of Patent: December 8, 1998) in view of Mantha et al. (US Patent No. 6,163,779 (US Patent No. 6,163,779, Date of Patent: December 19, 2000) and further in view of Finseth et al (US Patent No. 6,271,840, Date of Patent: August 7, 2001).

(CURRENTLY AMENDED) Claims 8 and 10:

Regarding claims 8 and 10, Ishikawa teaches an information search method for crawling a web site via a network using a computer, said method comprising the steps of:

acquiring a web page as initial information and storing source code into a storage device (see Figure 4, all features, Ishikawa);

reading the source code of said web page from said storage device (see Figure 4, wherein text body is taking into account to be the source code prescribed in webpage, Ishikawa);

conducting a structure analysis of said web page (see Figure 4, wherein parent document list is conducting a structure analysis of web page, Ishikawa) and column 19, lines 4-7, wherein user can realize document by reading the summary of each particular hypertext document without calling each particular hypertext document, Ishikawa), **wherein the structure analysis includes the steps of:**

Ishikawa discloses all the limitations above. However, Ishikawa is silent with respect wherein the source code is HTML nor does Ishikawa discloses reading an HTML document of a web page as an analyzing object.

On the other hand, Mantha discloses that the source code is HTML as well as **reading an HTML document of a web page as an analyzing object** (Figure 14, all features, Mantha). A skilled artisan would have been motivated to combine as suggested by Mantha, by clearly defining as well as stating that the source code is HTML, for creating hypertext documents for the World Wide Web, easy editing, and viewable documents.

conducting a temporary block analysis based on a description of HTML tags of the HTML document (column 1, lines 32-35 and column 2, lines 36-38, wherein the path name to the stored file is substituted for the original hypertext reference, wherein substitute is equivalent to temporary, Mantha),

using the HTML tags to temporarily divide the HTML document into blocks (column 19, lines 55-57, wherein a composition or an article is divided into a number of portions, and each portion of the composition is written in on of the hypertext documents, Ishikawa), **and**

identifying unnecessary information elements in the HTML document (column 11, lines 8-13, wherein this routines begins identifying the URL's original file and storing them in LL1, wherein preferably duplicate entries are eliminated, Mantha).

Ishikawa in view of Mantha discloses all the limitations above. However, Ishikawa in view of Mantha do not disclose wherein **the unnecessary information elements are plural information elements that include an OBJECT IMAGE having a same Uniform Resource Locator (URL)**; nor do they disclose **wherein the OBJECT IMAGE describes a type of media used to display the HTML**

document reading an HTML document of a web page as an analyzing object, On the other hand, Finseth discloses **wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL)** (see abstract, wherein search engine results or a list of URL's are passed to a web crawler that retrieves the web page and other media information present at the associated URL, wherein defined in applicants specification within the PG Publication, paragraph [0059], wherein OBJECT_IMAGE, represents all media types, Finseth), **and wherein the OBJECT_IMAGE describes a type of media used to display the HTML document** (column 5, lines 32-35 and lines 62-67, wherein rendering engines may include web document or HTML, Finseth). A skilled artisan would have been motivated to combine as suggested by Ishikawa in view of Mantha, removing duplicate information to provide the a user with relevancy and enabling faster perusal of search engine output as it relates to the world wide web.

storing a result of the analysis into said storing device (column 6,lines 31-35, 1shikawa);

calculating a degree of significance of a web site linking from said web page, based on the result of said structure analysis stored in said storage device (see Figure 5, all features and column 10, lines 18-25, Ishikawa); and

accessing the web site depending on the calculated degree of significance to acquire contents thereof (column 3, lines 4-19, Ishikawa), and storing them into said storage device (see Figure 3, diagram 8, Ishikawa).

Claim 11:

Regarding claim 11, Ishikawa in view of Mantha and further in view of Finseth teaches wherein said structure analyzing means associates mutually relevant information elements with each other, among information elements contained in said source code (see Figure 4, wherein word list is the relevant information elements and text body is the information elements contained in source code, and source code is considered to be taking into account the prescribed web page, and column 9, lines 4-16, Ishikawa).

Claim 12:

Regarding claim 12, Ishikawa in view of Mantha and further in view of Finseth teaches wherein said significance calculating means selects plural strategies as strategies for calculating the degree of significance of (column 10, lines 50-61, wherein selecting a plurality of candidates for a key word, Ishikawa) said web site (column 10, lines 41-46, wherein the user calls the HTML document in the world wide web, Ishikawa), and uses them by giving weights thereto, respectively (column 11, lines 47-65, wherein weights are assigned he number of keywords is two or more, it is applicable that an estimated value for one particular hypertext document be set to a value N times, i.e. N is two or more as high as a sum of the products TF*IDF calculated for all keywords when N particular words agreeing with N keywords appear in the particular hypertext document and . where two particular words agreeing with two keywords are used in one particular hypertext document close to each other within 20 characters, it is applicable that an estimated value for the unified particular hypertext document be doubled, Ishikawa).

(NEW) Claim 15:

Regarding claim 15, Ishikawa in view of Mantha and further in view of Finseth teaches reading an HTML document of a web page as an analyzing object (**REFER** to claim 8, wherein this limitation has already been addressed, Mantha);

conducting a temporary block analysis based on a description of HTML tags of the HTML document (**REFER** to claim 8, wherein this limitation has already been addressed, Mantha);

using the HTML tags to temporarily divide the HTML document into blocks (**REFER** to claim 8, wherein this limitation has already been addressed, Ishikawa);

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document (**REFER** to claim 8, wherein this limitation is substantially the same/similar to the limitation defined within claim 8, Mantha); and

deleting any block in the HTML document that is deemed to be structurally meaningless, wherein a block is deemed to be structurally meaningless if that block has only unnecessary information elements

(columns 8-9, lines 65-67 and line 1-6, wherein removing from a plurality of character strings existing in a body of collected reference document to form a text body, Ishikawa);

merging relevant information elements in a same block into one composite element (**REFER** to claim 13, wherein this limitation has already been addressed, Ishikawa).

(NEW) Claim 16:

Regarding claim 16, Ishikawa in view of Mantha and further in view of Finseth teaches wherein the unnecessary information elements include OBJECT_ANCHORS that have a same title, wherein an OBJECT_ANCHOR describes a correlation between the HTML document and elements in another web page (column 9, lines 52-64 and column 10, lines 48-63, Finseth).

(NEW) Claim 17:

Regarding claim 17, Ishikawa in view of Mantha and further in view of Finseth teaches wherein the unnecessary information elements include OBJECT_TEXT_BLOCKS that have the same description of text in a block (Figures 6 and 7, all features, further define in column 9, lines 52-55, wherein an image map is associated with the rendered image then clicking upon different areas of the rendered image will have the same results as those set forth in the description of Figure 6, Finseth).

(NEW) Claim 18:

Regarding claim 18, teaches wherein the relevant information elements that are merged are from a group that includes the OBJECT_IMAGE, OBJECT_ANCHOR and OBJECT_TEXT_BLOCKS (Figure 4, all features, Ishikawa).

(NEW) Claim 19:

Regarding claim 19, Ishikawa in view of Mantha and further in view of Finseth teaches wherein a method for eliminating ambiguity of a specified topic being searched during a web crawling, the method comprising:

presenting relevant keywords to a user during web crawling, wherein the relevant keywords describe multiple attributes of a term that has an ambiguous meaning, and wherein the user is afforded an ability to specify keywords that have a minus degree of significance to a meaning intended by the user for web crawling (column 1, lines 53-67, Ishikawa, wherein a plurality of identification titles indicating the

retrieval documents are arranged in increasing order, wherein increasing is equivalent to minus, when the user selects the identification titles displayed on the unit one after another in arranged order the retrieval document indicated by the selected identification title is read out from the document, wherein the title is equivalent to a keyword, Ishikawa); and

narrowing down crawling objects by eliminating user-specified keywords that have a minus degree of significance, thereby eliminating ambiguity of a term being searched (columns 8-9, lines 65-67 and lines 1-6 and lines 13-16, wherein removing is equivalent to deleting, from a plurality of strings existing in a body of collected reference document to form a text body, and the text body is written in the document information entry space, and wherein a plurality of words used in the text body the title and the anchor sentence are written in the document, wherein words is equivalent to term, Ishikawa).

(NEW) Claim 20:

Regarding claim 20, Ishikawa in view of Mantha and further in view of Finseth teaches a web crawler comprising:

an initial site acquiring section, wherein the initial site acquiring section specifies a Uniform Resource Locator (URL) of a home page of a specific web site from which information is to be collected, (Figure 4, diagrams 31-36 Mantha) and wherein initial web sites to be searched are obtained through the use of keywords in a search engine (Figure 5C, all features, wherein its further define in column 8, lines 40-54, Mantha), and wherein the initial web sites that are initially set for web crawling (Figure 5B, all features, wherein yahoo is the initial page, and its known to a web crawling site, Mantha);

a document structure analysis section for performing document structure analysis for a web page of initial sites, wherein the document structure analysis includes the step of:

reading an HTML document of a web page as an analyzing object (**REFER** to claim 8, wherein this limitation has already been addressed);

conducting a temporary block analysis based on a description of HTML tags of the HTML document (**REFER** to claim 8, wherein this limitation has already been addressed);

using the HTML tags to temporarily divide the HTML document into blocks (**REFER** to claim 8, wherein this limitation has already been addressed);

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document (**REFER** to claim 8, wherein this limitation has already been addressed); and

a significance calculating section for calculating degrees of significance of web sites that are acquired by web crawling, wherein the degrees of significance are based on a result of the document structure analysis performed by the document structure analysis section (column 4, lines 22-39 and column 10, lines 18-26, Ishikawa), and wherein calculating degrees of significance extends a Fish-search crawling technique by basin the calculating on strategies specified by a user and information elements added to anchors through the document structure analysis section, and wherein objects of crawling are dynamically determined depending on the degree of significance (column 17, lines 51-62, Ishikawa); and

a crawling executing section for executing a process of acquiring web sites by crawling based on the results of the degrees of significance calculated by the significance calculating section (column 4, lines 35-40, Ishikawa).

Prior Art of Record

1. Ishikawa et al (US Patent No. 5,848,407) discloses a hypertext document and anchor sentences of parent documents for the hypertext document are registered with an hypertext document identifier as document information for each of hypertext documents having reference relationships with each other.
2. Finseth et al (US Patent No. 6,271,840) discloses a visual index method provides graphical output from search engine results or other URL lists.
3. Mantha et al (US Patent No. 6,163,779) discloses a method of copying a web page presented for display on a browser of a Web client.

Point of Contact

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Helene Rose whose telephone number is (571) 272-0749. The examiner can normally be reached on 8:00am - 4:30pm Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Don Wong can be reached on (571) 272-1834. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

HRR
Technology Center 2100
September 1, 2006

DON WONG
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100